DISTRIBUTED SOURCE-CHANNEL CODING FOR WIRELESS SENSOR NETWORKS

Michael Gastpar

University of California, Berkeley
Department of EECS
Berkeley, CA 94720-1770
gastpar@eecs.berkeley.edu

ABSTRACT

In this paper, we investigate properties of good coding strategies for a class of wireless sensor networks that could be termed "monitoring" networks: Their task is to monitor an underlying physical reality at the highest possible fidelity. Since the sensed signals are often analog, and the communication channels noisy, it will not generally be possible to exactly communicate the sensed signals. Rather, such sensor network scenarios involve both a *compression* and a *communication* problem. It is well known that these two tasks must be addressed jointly for optimal performance, but optimal performance is unknown in general.

This problem is addressed from a scaling-law perspective in this paper, i.e., as the number of nodes becomes large. The goal of the paper is to characterize the key properties of coding strategies that achieve the optimum scaling behavior, and hence to identify the scaling-law relevant issues in code design. We first present a lower bound to the cost-distortion tradeoff, and then compare two fundamentally different coding strategies to that lower bound.

1. INTRODUCTION

The class of sensor networks of interest to this study could be termed *monitoring sensor networks:* Their goal is to observe a physical system over time and space at the highest possible fidelity. A simple example of such a sensor network was analyzed in [1], and generalizations thereof in [2]. The present paper further investigates the conditions under which a simple uncoded forwarding strategy attains the optimal scaling behavior, and characterizes the properties of scaling-law optimal codes for certain scenarios.

The sensor network model studied in this paper is shown in Figure 1. There is a physical phenomenon, characterized by L variables, representing the degrees of freedom of the system, or, equivalently, its current state. We model each degree of freedom as a random process in discrete time.[1]

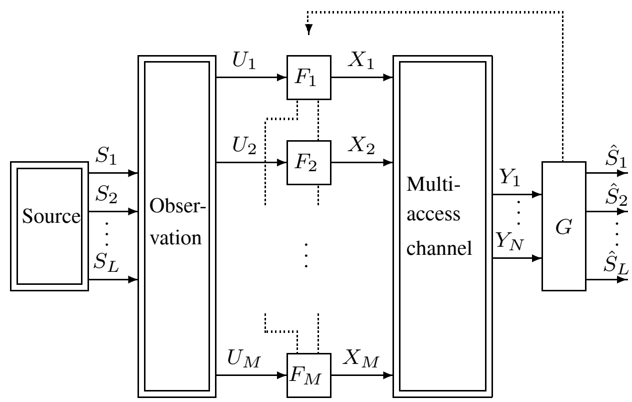

Fig. 1. The "monitoring" sensor network topology considered in this paper.

The underlying L sources are observed through a noisy observation mechanism by M sensors. As expressed by the dotted lines in Figure 1, the sensors may have the possibility to collaborate to some (generally limited) extent, and there may be feedback from the base stations to each of the sensors. Based on the respective sensor readings, the inter-sensor communication, and the feedback signals, the sensors communicate to N base stations, for example across a wireless medium. We assume the N base stations to be ideally connected through separate links (e.g., wireless, but over a different frequency band; or a fiber-optical link). The parameter N, in an appropriate sense, models the spatial channel bandwidth. For each source sample, the sensors can use K channel uses (modeling the temporal channel bandwidth), and a total power (or, more generally, *cost*) of P_{tot}. The cost is measured at the channel inputs and takes the shape of an expectation,

$$E\rho(X_1, X_2, \ldots, X_M) \quad \leq \quad P_{tot}. \tag{1}$$

In some related applications, there may be *multiple* cost

[1] The discrete-time model is justified by arguing that the state of the system does not change very rapidly. This may be a serious restriction for certain scenarios. The continuous-time extension is currently under investigation.

constraints, modeling the requirement of separate power limits for each sensor, rather than a sum power constraint. Eventually, the underlying sources can be reconstructed at a distortion D, with takes the shape of an expectation,

$$D = Ed\left((S_1, \ldots, S_L), (\hat{S}_1, \ldots, \hat{S}_L)\right). \quad (2)$$

for an appropriately chosen (application specific) distortion measure $d(\cdot, \cdot)$.

Our goal is to characterize the fundamental relationships between the six entities, i.e., L, N, K, M, P_{tot}, and D, with particular emphasis on the case where the number of sensing terminals, M, becomes very large. In a second stage, we also discuss some of the properties of coding strategies that achieve optimal relationships.

Scaling Law Notation

In this paper, we establish *scaling laws*, denoted by the symbol $\sim$, which here is taken to mean "asymptotic equivalence." More precisely, we write scaling laws as

$$f_1(M) \sim f_2(M), \quad (3)$$

which simply means that $\lim_{M\to\infty} f_1(M)/f_2(M) = c$, for some constant $c > 0$. The special case when $c = 1$ will be called a *strong scaling law*, since it correctly reports *both* the scaling behavior *and* the important constants, and will be denoted as $f_1(M) \overset{=}{\sim} f_2(M)$.

2. GAUSSIAN SENSOR NETWORKS

Of particular interest to the arguments of this paper is a special case of the sensor network of Figure 1, namely when all involved statistics are Gaussian. The resulting scenario is illustrated in Figure 2. In particular, there are L physical sources (the spatial or temporal bandwidth of the source), M sensors, and N receivers (base stations). The receivers are assumed to be ideally linked to each other: in the considered network model, the data collection point has access to the exact received value at each of the N base stations.

2.1. Network Parameters

Source Bandwidth L and Observation Process
The source is characterized by L independent and identically distributed (iid)[2] circularly complex Gaussian random variables with mean zero and variance σ_S^2. The observation process is modeled as

$$U_m = W_m + \sum_{l=1}^{L} a_{m,l} S_l, \quad (4)$$

[2]Assuming iid sources is without loss of generality in the sense that the matrix A in Eqn. (5) below can be chosen arbitrarily.

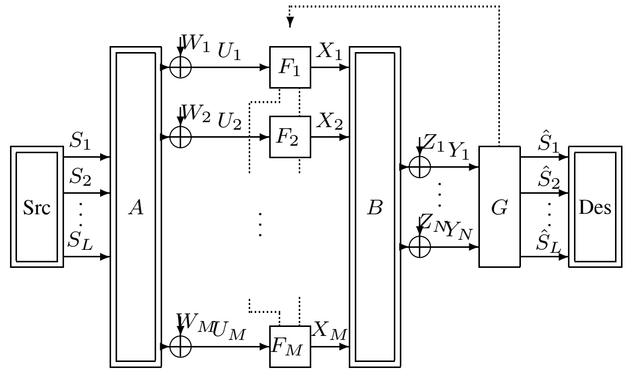

Fig. 2. The considered Gaussian sensor network.

where W_m is iid circularly complex Gaussian with mean zero and variance σ_W^2. We collect the coefficients $a_{m,l}$ into the matrix $A \in \mathbb{C}^{M\times L}$ defined as

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,L} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,L} \\ a_{3,1} & a_{3,2} & \cdots & a_{3,L} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M,1} & a_{M,2} & \cdots & a_{M,L} \end{pmatrix} \quad (5)$$

The matrix A has $\min\{M, L\}$ singular values that we denote by $\alpha_1, \alpha_2, \ldots, \alpha_{\min\{M,L\}}$. The parameter $\mathrm{rank}(A)$ models the product of the spatial and temporal bandwidth of the underlying source.

For the scope of the present paper, the coefficients $a_{m,l}$ are chosen according to a given distribution, and we suppose that their values are known throughout the network. Later, we argue that under certain circumstances, the sensors need not know these values, and the destination only needs limited knowledge, without changing the scaling behavior.

Spatial Bandwidth N of the Communication Channel
The communication channel is the standard additive white Gaussian multiple access channel, modeled as

$$Y_n = Z_n + \sum_{m=1}^{M} b_{n,m} X_m, \quad (6)$$

where Z_n is iid circularly complex Gaussian with mean zero and variance σ_Z^2. For the purpose of the present paper, the coefficients $b_{n,m}$ are fixed and assumed to be known throughout the network.[3] We collect the coefficients $b_{n,m}$ into the matrix $B \in \mathbb{C}^{N\times M}$ defined as

$$B = \begin{pmatrix} b_{1,1} & b_{1,2} & b_{1,3} & \cdots & b_{1,M} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{N,1} & b_{N,2} & b_{N,3} & \cdots & b_{N,M} \end{pmatrix} \quad (7)$$

[3]In [3], we discuss the case where the coefficients $b_{n,m}$ are randomly chosen, and unknown to the sensors, hence modeling the situation of limited synchronization between the sensors.

The matrix B has $\min\{N, M\}$ singular values that we denote by $\beta_1, \beta_2, \ldots$. Our interest is in the case where $N \leq M$, and in this paper, we consider the case where the matrix B has full rank. Hence, N models the spatial bandwidth of the communication channel.

Temporal Bandwidth K of the Communication Channel
The channel can be used K times for each source sample. This is equivalent to multiplying the bandwidth of the channel by a factor of K, and hence permits to study the temporal bandwidth of the channel.

Power on the Communication Channel
The power on the communication channel is constrained as

$$\sum_{m=1}^{M}\sum_{k=1}^{K} E|X_{m,k}|^2 \leq P_{tot}, \tag{8}$$

i.e., P_{tot} denotes the total power available per source output $(S_1, S_2, \ldots, S_L)$.

Target Distortion
The goal of the sensor network is to minimize the mean-squared error,

$$D = \frac{1}{L}\sum_{l=1}^{L} E|S_l - \hat{S}_l|^2, \tag{9}$$

where the expectation is over the distribution of the source vector S, the distribution of all the noises $W_1, \ldots, W_M, Z$, as well as the distribution of the observation matrix A.

2.2. Lower Bound to the Optimal Distortion Scaling

In [3], we present a lower bound to the distortion that can be achieved in the Gaussian sensor network defined in Subsection 2 for fixed observation matrix A. This lower bound can be adapted to the case of a randomly chosen matrix A. For the case of sufficiently large total power P_{tot}, it takes the following simple shape:

Theorem 1. *The distortion that can be achieved in the Gaussian sensor network defined in Subsection 2, under the assumption that the total sensor power P_{tot} is large enough (see Remark 2), cannot be smaller than*

$$\begin{aligned} D_{lower}&(M, P_{tot}, L, K, N) \\ &= E_A\left[\frac{1}{L}\sum_{l=1}^{L}\frac{\sigma_S^2\sigma_W^2}{\alpha_l^2\sigma_S^2+\sigma_W^2}\right] \\ &\quad +\nu\left(\frac{1}{\mu+\frac{P_{tot}}{KN\sigma_Z^2}\sqrt[N]{\prod_{n=1}^{N}\beta_n^2}}\right)^{KN/L}, \end{aligned} \tag{10}$$

where σ_S^2 is the variance of the underlying sources, σ_W^2 is the variance of the observation noises, σ_Z^2 is the variance of the noise in the multi-access channel, P_{tot} is the total sensor transmit power for the K channel uses, N is the number of destination terminals, and

$$\mu = \left(\frac{1}{N}\sum_{n=1}^{N}\frac{1}{\beta_n^2}\right)\sqrt[N]{\prod_{n=1}^{N}\beta_n^2} \tag{11}$$

$$\nu = E_A\left[\sqrt[L]{\prod_{l=1}^{L}\frac{\alpha_l^2\sigma_S^4}{\alpha_l^2\sigma_S^2+\sigma_W^2}}\right]. \tag{12}$$

The proof of this theorem is given in [3].

Remark 1. This outer bound includes the case of arbitrary collaboration between the sensors, and of arbitrary feedback signals from the data collection point to the sensors.

Remark 2. In Theorem 1, we assume that P_{tot} is "large enough." For small P_{tot}, the solution involves inverse water-filling, i.e., not all of the eigenvalues will be encoded. We will provide details in [3].

3. ACHIEVABLE SCALING BEHAVIOR

3.1. Separate Source and Channel Coding

It is well known that separate source and channel coding does not lead to optimal performance in general networks, but it is generally hard if not impossible to determine the best separation-based performance. Here, we give a lower bound to the corresponding distortion scaling law, combining the rate-distortion results of [5, 6, 7] for the so-called CEO problem with the capacity of the Gaussian multiple-input multiple-output channel with inputs $X_1, \ldots, X_M$ and outputs $Y_1, \ldots Y_n$, see [4]. To make matters simple, we suppose that A and B are fixed matrices, that the coefficients of the matrix A all have unit magnitude, and that the rows of the matrix A are orthogonal. For sufficiently large M, it takes the following simple shape:

$$\begin{aligned} D&(M, P_{tot}, L, K, N) \\ &= \frac{\sigma_S^2\sigma_W^2}{M\sigma_S^2+\sigma_W^2}+\frac{\sigma_S^4}{\sigma_S^2+\sigma_W^2/M} \\ &\quad \cdot \frac{1}{1+\frac{KN\sigma_S^2}{L\sigma_W^2}\log_2\left(\mu+\frac{P_{tot}\beta_0^2}{KN\sigma_Z^2}\sqrt[N]{\prod_{n=1}^{N}\beta_n^2}\right)} \end{aligned} \tag{13}$$

where μ is given by (11). In the next paragraph, we consider a simple *joint* source-channel coding technique. A discussion and comparison of the two approaches following in Subsection 3.3.

3.2. Joint Source-Channel Coding Techniques

In [2], we consider a simple joint source-channel coding strategy, whose essence is for each sensor to transmit a scaled version of its observation. This simple analysis extends directly to the scenario of Figure 2 whenever the matrices A and B are "matched" in an appropriate fashion. The simplest notion of matched matrices in this context is to require $N = L$ and

$$BA \quad = \quad \operatorname{diag}(\lambda_1, \ldots, \lambda_L), \tag{14}$$

where $\operatorname{diag}(x_1, \ldots, x_L)$ denotes a diagonal $L \times L$ matrix with diagonal entries $x_1, \ldots, x_M$. More general notions will be presented in [3].

For matched matrices, the scaling law (10) is indeed achievable. Under the same hypothesis on the entries of the matrices A and B as in Section 3.1, Equation (10) becomes

$$\begin{aligned} & D_{lower}(M, P_{tot}, L = N, K = 1) \\ = & \quad \frac{\sigma_S^2 \sigma_W^2}{M\sigma_S^2 + \sigma_W^2} + \frac{\sigma_S^4}{\sigma_S^2 + \sigma_W^2/M} \\ & \quad \cdot \frac{1}{\mu + \frac{P_{tot}}{N\sigma_Z^2} \sqrt[N]{\prod_{n=1}^N \beta_n^2}}. \end{aligned} \tag{15}$$

This result is easily extended to the more general case $L = KN$.

3.3. Comparison

Comparing (13) with (15), there is a key difference in the respective second terms. To interpret its significance, consider the following short argument. In any case, the best decay of the distortion as a function of the number of nodes is

$$D(M) \quad \sim \quad \frac{1}{M}. \tag{16}$$

Hence, the key question becomes that of how much resources must be allocated to actually harvest this optimal distortion scaling behavior. This is tantamount to requiring the second terms in (13) and (15) to *also* decay like $1/M$.

To gain insight into this issue, let us suppose for now that L, K, and N are all fixed, and that M becomes very large. For (13), this then amounts to requiring that

$$P_{tot,sep} \quad \sim \quad e^M, \tag{17}$$

i.e, that the total power $P_{tot,sep}$ increase *exponentially* in M - this is the only way to make the second term in (13) decay like $1/M$. By contrast, for joint source and channel coding under the assumptions of Subsection 3.2, a total power that satisfies

$$P_{tot,joint} \quad \sim \quad M, \tag{18}$$

i.e.,that increases *linearly* as a function of the number of nodes is sufficient. Moreover, in many cases of interest, the values β_n^2 increase linearly in M. For those cases, a *constant* total power is already sufficient.

4. CONCLUSIONS AND EXTENSIONS

In this paper, our focus is on identifying the scaling-law relevant characteristics of coding schemes for sensor networks of the type illustrated in Figure 1. We have shown by way of an example that in general, strategies in which each sensor tries to get his own bits to the data collector without making errors are doomed to achieve an *exponentially* suboptimal scaling behavior. Instead, strategies must be designed taking into account the source and channel structures. For our example, we have seen that source and channel structures are already matched, and (almost) uncoded transmission by the sensors achieves a scaling-law optimal performance. In more general cases, the sensors must attempt to match the source structure to the channel structure. Approaches on how this could be done are along the lines of [8]. This will be presented in [3].

Acknowledgment. This work is an extension of joint work of the author with Prof. Martin Vetterli (EPFL).

5. REFERENCES

[1] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," in *2nd International Workshop on Information Processing in Sensor Networks (IPSN'03)*, Leonidas J. Guibas and Feng Zhao, Eds., pp. 162–177. Lecture Notes in Computer Science, Springer, New York, NY, April 2003.

[2] M. Gastpar and M. Vetterli, "Scaling laws for homogeneous sensor networks," in *Proc 2003 Allerton Conference*, October 2003, Invited paper.

[3] M. Gastpar and M. Vetterli, "Scaling laws for sensor networks," *to be submitted to IEEE Transactions on Information Theory*, November 2003.

[4] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Bell Labs Technical Memorandum*, June 1995, Also published in *European Transactions on Telecommunications*, 10(6):585-596, Nov.-Dec. 1999.

[5] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Transactions on Information Theory*, vol. IT–42, pp. 887–902, May 1996.

[6] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Transactions on Information Theory*, vol. IT–44, no. 3, pp. 1057–1070, May 1998.

[7] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Transactions on Information Theory*, vol. IT–43, no. 5, pp. 1549–1559, September 1997.

[8] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.